Introduction to USDA Integrated Pathogen Modeling Program – Global Fit

(IPMP-Global Fit)

Lihan Huang, Ph.D.

Residue Chemistry and Predictive Microbiology Research Unit

Eastern Regional Research Center

USDA Agricultural Research Service

600 E. Mermaid Lane

Wyndmoor, PA 19038

Edition: January, 2017

Contents

DISCLAIMER AND ASSUMPTION OF RISK

The USDA Integrated Pathogen Modeling Program – Global Fit (IPMP Global Fit) is a software tool developed by the USDA Agricultural Research Service (ARS) for data analysis and model development in predictive microbiology. USDA grants to each recipient of this software non-exclusive, royalty free, world-wide, permission to use, copy, publish, distribute, perform publicly and display publicly this software. We would appreciate acknowledgement if the software is used.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT AND ANY WARRANTY THAT THIS SOFTWARE IS FREE FROM DEFECTS. IN NO EVENT SHALL USDA BE LIABLE FOR ANY CLAIM, LOSS, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

The risk of any and all loss, damage, or unsatisfactory performance of this software rests with you, the recipient. USDA provides no warranties, either express or implied, regarding the appropriateness of the use, output, or results of the use of the software in terms of its correctness, accuracy, reliability, being current or otherwise. USDA has no obligation to correct errors, make changes, support this software, distribute updates, or provide notification of any error or defect, known or unknown. If you, the recipient, rely upon this software, you do so at your own risk and you assume the responsibility for the results. Should this software prove defective, you assume the cost of all losses, including but not limited to, any necessary servicing, repair or correction of any property involved.

Please contact Dr. Lihan Huang (Lihan.Huang@ars.usda.gov) for technical questions.

SUGGESTED CITATION

Huang, L. 2017. USDA Integrated Pathogen Modeling Program – Global Fit (http://www.ars.usda.gov/Main/docs.htm?docid=23355). USDA Agricultural Research Service, Eastern Regional Research Center, Wyndmoor, PA.

INTRODUCTION

What is IPMP - Global Fit?

IPMP-Global Fit is an extension of the USDA Integrated Pathogen Modeling Program (IPMP). It is designed as a one-step direct kinetic analysis tool that constructs a tertiary model from the entire experimental data set of growth and inactivation. IPMP-Global Fit tries to analyze and fit the entire experimental data for both primary and secondary models and obtain the kinetic parameters that minimize the global error of the primary model. It differs from IPMP, which is design to analyze individual growth or inactivation curves. IPMP-Global Fit will analyze all isothermal growth or inactivation curves obtained under different conditions together, and try to derive kinetic parameters that are optimized for the entire data set. It is a tool developed to implement the one-step kinetic analysis methodology proposed by Huang (2015a, 2015b, and 2016).

Why IPMP-Global Fit?

The development of primary models in predictive microbiology usually involves a two-step process. The first step is to analyze each individual growth or inactivation curve obtained under a constant environmental condition (temperature, for example). The kinetic parameters (growth rate or lag time) from different conditions are then used to derive a secondary model.

IPMP-Global Fit takes a different approach. It implements the one-step kinetic analysis methodology (Huang, 2015a, 2015b, and 2016). It is designed to simultaneously analyze the entire data set in search for the best estimates of the kinetic parameters for both primary and secondary models. IPMP-Global Fit tries to minimize the global error between the data and the primary model, producing estimates for each parameter for both the primary and secondary models. From one-step kinetic analysis, it is possible to directly estimate and interpret the errors of the primary model.

What is difference between IPMP and IPMP-Global Fit?

IPMP analyzes each individual curves, but IPMP-Global Fit works on the entire data from the entire experiment.

What is required to use IPMP-Global Fit?

All the statistical analysis and model development are handled seamlessly behind the scenes. No programming knowledge is needed. The users only need to enter the data and click a few buttons on the screen to complete any data analysis. The only requirement is that the users have a basic knowledge of predictive microbiology to allow for the selection of suitable models for data analysis.

What models are included in IPMP-Global Fit?

The development of IPMP-Global Fit was originally inspired by the need to solve for a common exponent of different curves of the Weibull model. It has been expended to include both growth and survival models, including the Weibull model, 1st thermal inactivation model, no-lag phase model, and Huang Model. For 1st thermal inactivation model, IPMP will determine a global log(D_0) and z value. For growth, IPMP Global Fit allows you to analyze the data with different secondary models, including Ratkowsky square-root model, Huang square-root model, and cardinal parameters model.

STRUCTURE of IPMP-Global Fit

IPMP-Global is based on IPMP. So they share similar components and design. If you are familiar with IPMP, you should be able to use IPMP-Global Fit. The major difference is in the data window, which is explained in the Data Window Section (Figure 1).

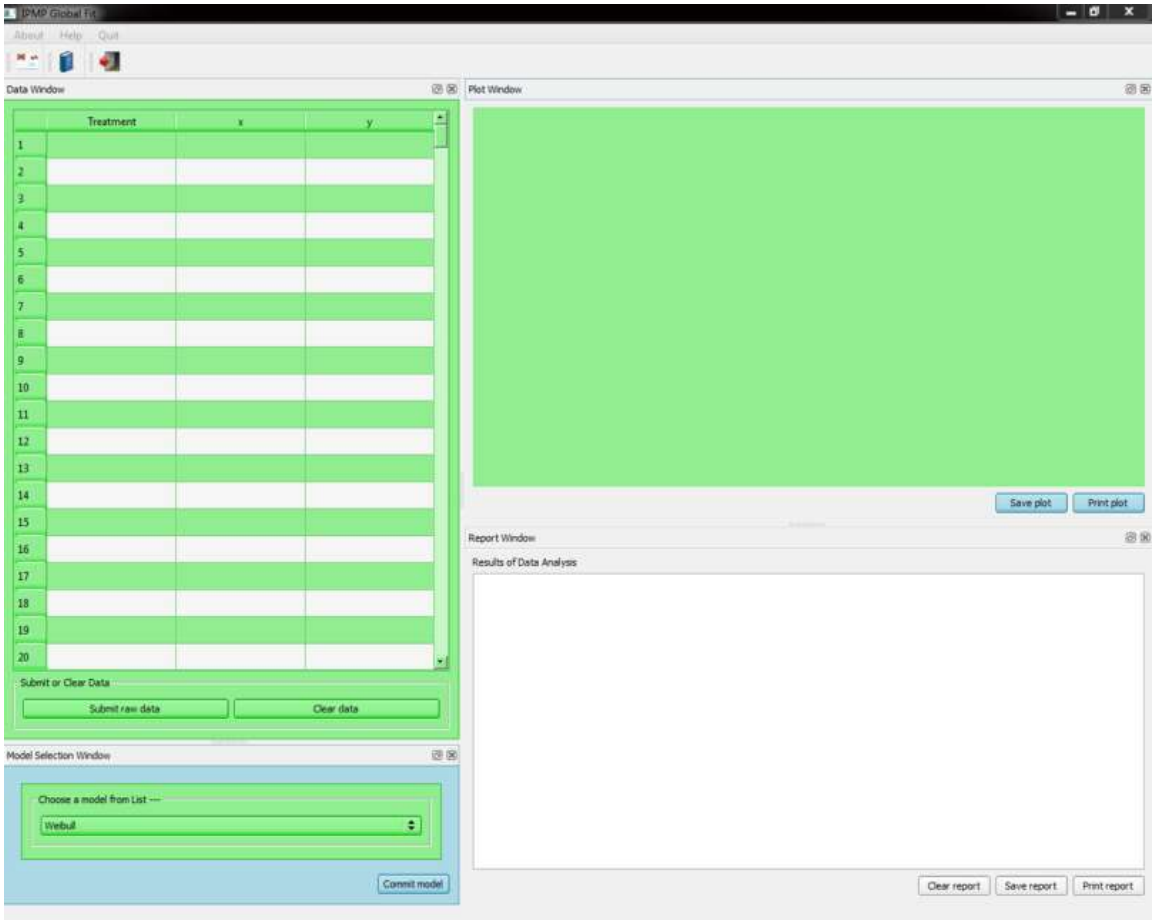

Figure 1. Four windows in IPMP-Global Fit: Data Window, Plot Window, Model Selection, and Report Window.

DATA WINDOW

Components

The data window contains a spreadsheet-style input area (Figure 2). The data input area contains three columns and 5000 rows. The data table can be scrolled to examine the data.

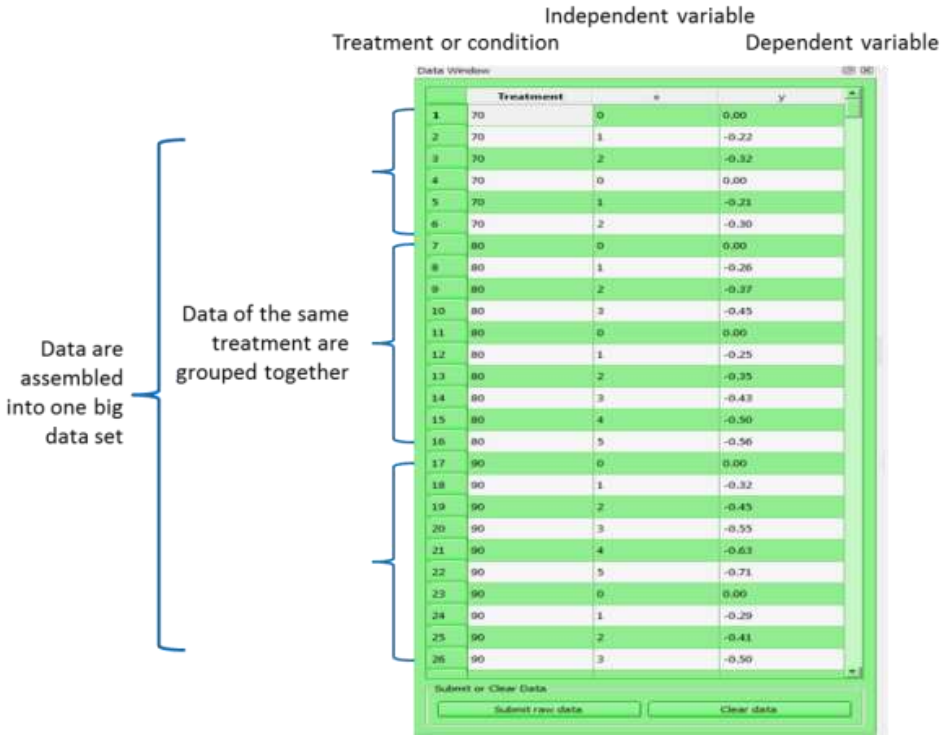

Figure 2. Data table.

Data format and assembly

The first column is reserved for "treatments" or "conditions" of experiment (temperature, for example). The secondary column is reserved for the independent variables (time, for example). The third column is reserved for dependent variable (y). It is recommended that the data of the same treatment from all replicates are grouped together (Figure 1) before assembling the entire data set.

Raw Data Entry

Raw data must be entered in the data input area. The raw data can be directly entered from the keyboard or copied/pasted from a text editor or a spreadsheet (Excel®, for example). The data can be growth or survival data. Raw data can be edited by right-clicking the mouse. The edit operations include "cut", "copy", "paste", and "clear". The data can be saved to "cvs" format by clicking the "save" option. If necessary, click "Clear data" to erase the data from the input area. Clear data before entering new data.

For thermal inactivation (Weibull and linear inactivation models), the y column must be in net growth, or negative of log reduction [$log(N) - log(N_0)$], where log(N) is the real time bacterial count, and log(N_0) is the initial bacterial count for each isothermal experiment. This transformation is needed for each survival curve. Data of the same treatment (temperature) can be grouped together.

For growth data, the y column must be in log cfu (not ln cfu) (Figure 2). All growth data will be converted to ln cfu internally by the software. The users must be aware that ln cfu is used in all primary growth models in IPMP Global Fit. The growth rate μ_{max} obtained from data analysis is specific growth rate in ln cfu h^{-1}.

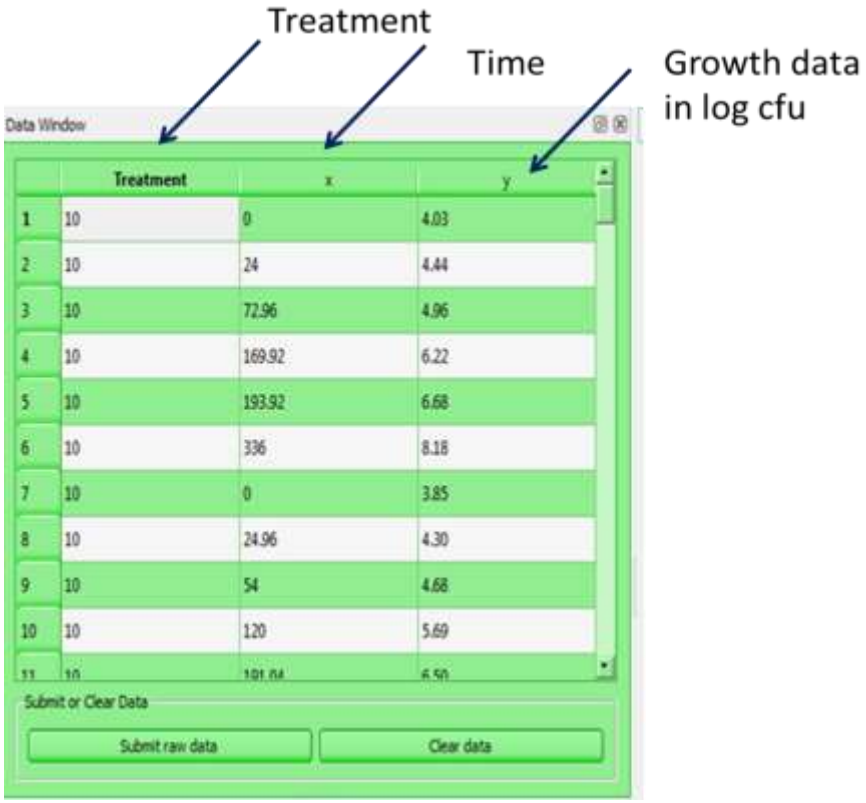

Figure 3. Data entry for growth curves.

Data Submission

Once the data are entered into the data window, click the "Submit Raw Data" button. All data will be automatically processed and sent to the plot window.

PLOT WINDOW

The Plot Window is used in IPMP-Global Fit for plotting data. All treatments will be plotted simultaneously. The Plot Window is also used in adjusting the parameters for growth curves. For growth curves, the users are expected to make preliminary selections of parameters and visually inspect the resulting plots. Once the analysis is completed, the regression curves will be plotted along with raw data.

MODEL SELECTION WINDOW

The model selection window contains four primary models, including two inactivation models and two growth models (Figure 4). For inactivation models (Weibull and D-z), the data will be automatically processed and analyzed. The regression results will be automatically plotted and reported. No further action from users is needed.

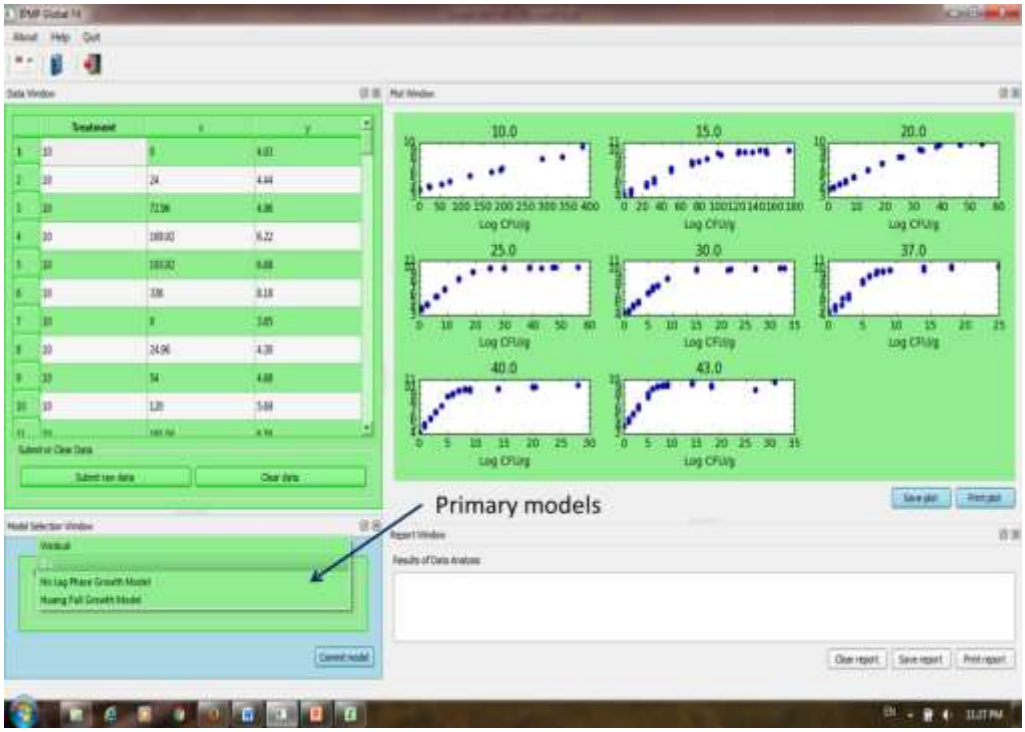

Figure 4. Selection of primary models.

For growth models, the users will need to choose from two models. The first is the No-Lag phase model, which is particularly suitable for growth curves with no lag phase or when the lag phase in the growth curves is negligible. For growth curves showing lag phases, the users would need to choose the Huang model. Once a primary growth model is chosen, click the “Commit Model” button. A pop-up window containing five secondary models will appear (Figure 5). The users are expected to select a secondary model and make adjustments of the parameters in the secondary model. Once a parameter is adjusted, the growth curves will be plotted along with the raw data (Figure 6). Once the users feels

that a good selection has been made, click the “Commit Model” button in the secondary model window. The data will be sent to the data engine for processing. If a convergence is found, the optimized curves will be plotted and results will the reported in the report window.

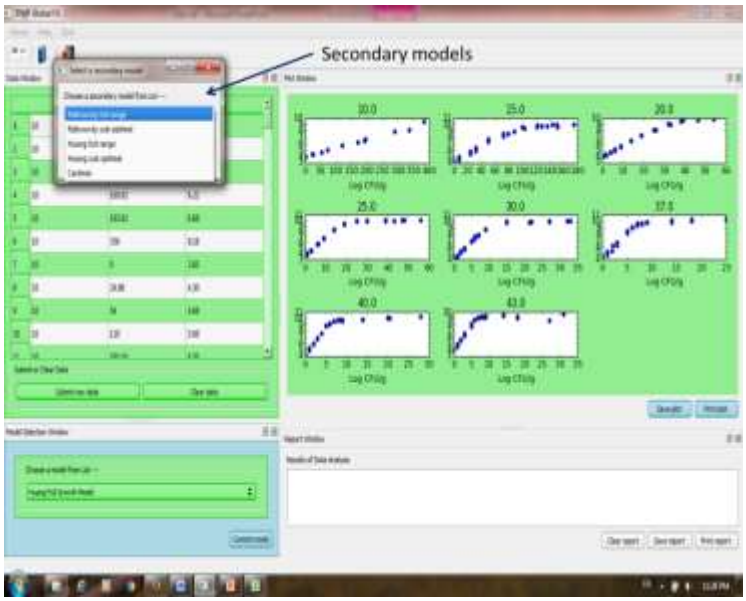

Figure 5. Selection of secondary models.

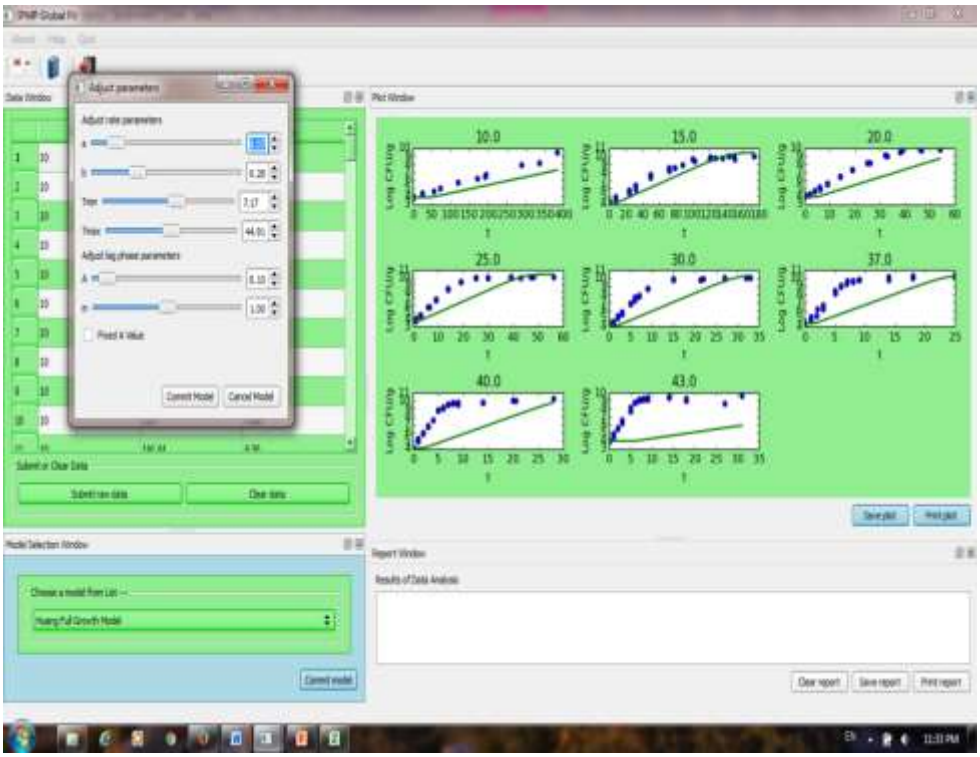

Figure 6. Adjustment of parameters in a secondary model.

REPORT WINDOW

If a convergence is found after data analysis, a report will be automatically generated (Figure 7).

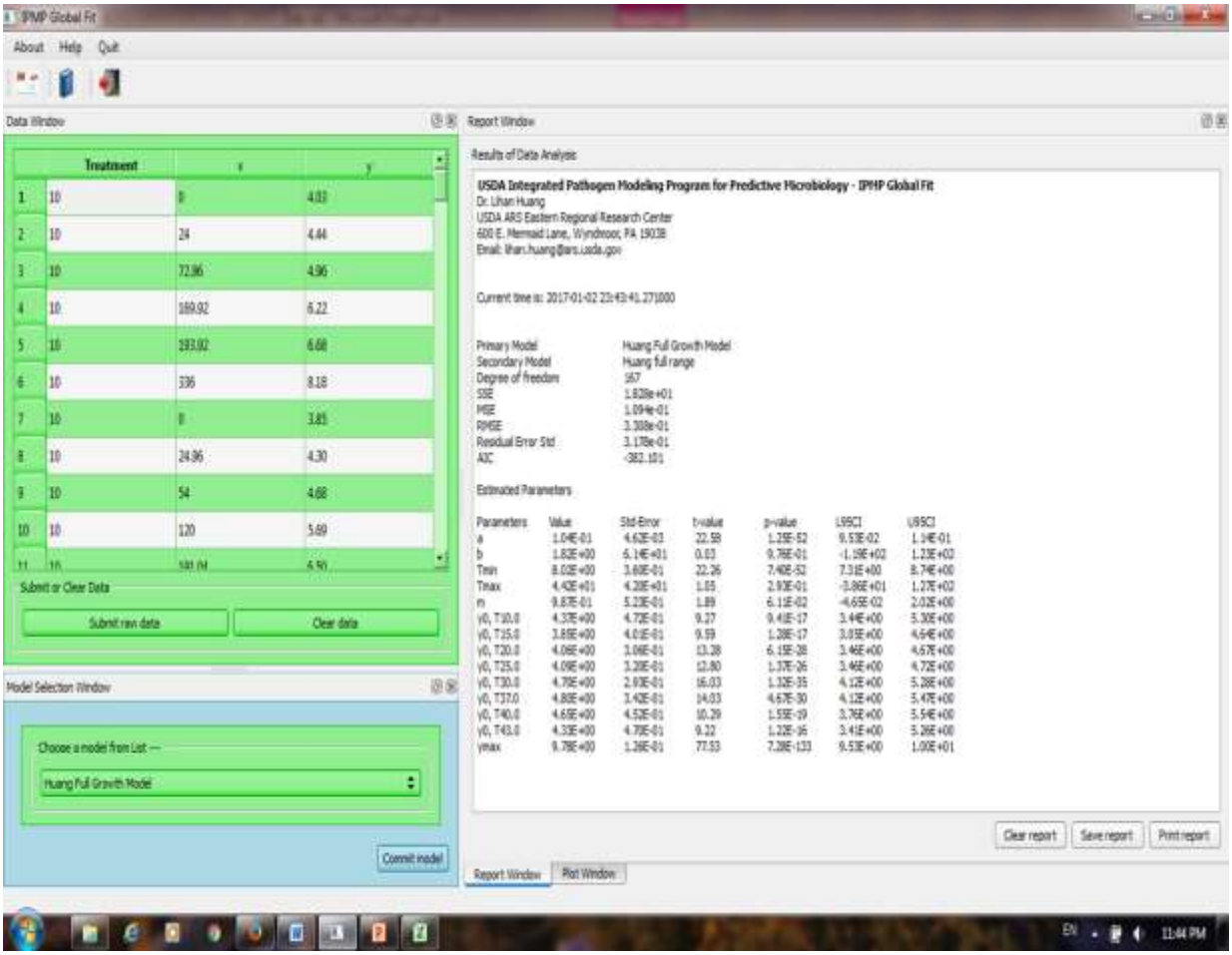

Figure 7. Data report window.

DATA ANALYSIS IN IPMP-GLOBAL FIT

The first step in data analysis to format and assemble the data. For inactivation kinetic analysis, the bacterial counts must be converted into net growth, or negative of log reduction. For growth, the growth data log cfu can be directly used. The data must be grouped according to treatment. The first column must be the treatment. The second column must be time. The third column must be the dependent variable (bacterial counts). The data can be formatted using Excel or a text editor.

The second step is to copy and paste the data into the Data Window, and then click the “Submit raw data” button. Data will be plotted automatically.

The third step is to make a selection in the Model Selection Window. For inactivation, select one of the inactivation models and he click "Commit model" button. For growth model, select either one of the two growth models, and then click "Commit model". A new pop-up window will appear for the users to choose a secondary model. Choose a secondary model and adjust the parameters to observe the plots. Once a reasonable choice is made, click the "Commit Model" button in the pop-up window. The data will be processed and analyzed automatically.

MATHEMATICAL MODELS IN IPMP-Global Fit

1 Thermal inactivation models

1.1 Nonlinear Weibull- Mafart model (Mafart, Couvert, Gaillard, and Leguerinel, 2002)

$$\boldsymbol{log(N) - log(N_0) = y_0 - \left(\frac{t}{D}\right)^K}$$

The objective of IPMP-Global Fit is to find a common K for the entire data set, and different D for each temperature condition. To use IPMP-Global Fit for Weibull model, the dependent variable is log reduction, or $log(N) - log(N_0)$. This transformation is needed for each survival curve. Data of the same treatment (temperature) can be group together. log(N) is the real time bacterial count, and $log(N_0)$ is the initial bacterial count for each isothermal experiment. There is no secondary model for the Weibull model.

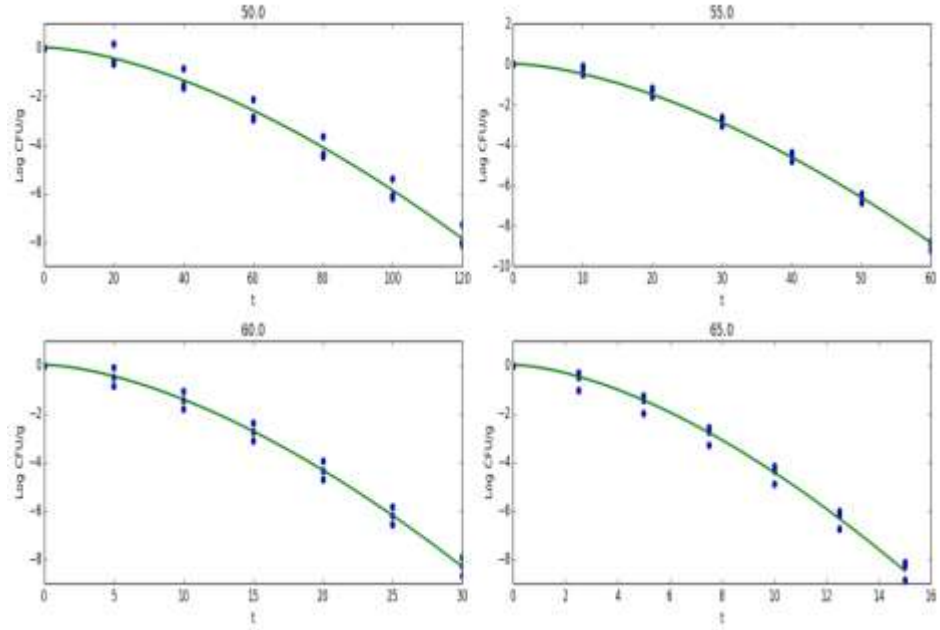

Figure 8. Multiple survival curves are analyzed simultaneously using the Nonlinear Weibull- Mafart model.

USDA Integrated Pathogen Modeling Program for Predictive Microbiology - IPMP Global Fit

Dr. Lihan Huang
USDA ARS Eastern Regional Research Center
600 E. Mermaid Lane, Wyndmoor, PA 19038
Email: lihan.huang@ars.usda.gov

Current time is: 2017-01-03 15:25:10.010000

Primary Model	Weibull
Secondary Model	None
Degree of freedom	78
SSE	6.962e+00
MSE	8.925e-02
RMSE	2.988e-01
Residual Error Std	2.879e-01
AIC	-193.717

Estimated Parameters

Parameters	Value	Std-Error	t-value	p-value	L95CI	U95CI
D, T50.0	3.31E+01	4.26E+00	7.76	2.75E-11	2.46E+01	4.16E+01
D, T55.0	1.54E+01	2.03E+00	7.56	6.55E-11	1.13E+01	1.94E+01
D, T60.0	8.00E+00	1.04E+00	7.67	4.06E-11	5.92E+00	1.01E+01
D, T65.0	3.94E+00	5.17E-01	7.63	4.78E-11	2.92E+00	4.97E+00
K	1.60E+00	1.48E-01	10.80	3.79E-17	1.31E+00	1.90E+00
y0	2.11E-02	2.38E-01	0.09	9.30E-01	-4.53E-01	4.96E-01

Figure 9. Results of data analysis for the Weibull model.

1.2 Linear inactivation model: D-z model

1.2.1 Primary model

$$\boldsymbol{log(N) - log(N_0) = y_0 - \frac{t}{D}}$$

1.2.2 Secondary model

$$\boldsymbol{log(D) = log(D_0) - \frac{T}{D}}$$

To use IPMP-Global Fit for D-z model, the dependent variable is log reduction, or $log(N) - log(N_0)$. This transformation is needed for each survival curve. Data of the same treatment (temperature) can be group together. log(N) is the real time bacterial count, and log(N_0) is the initial bacterial count for each isothermal experiment.

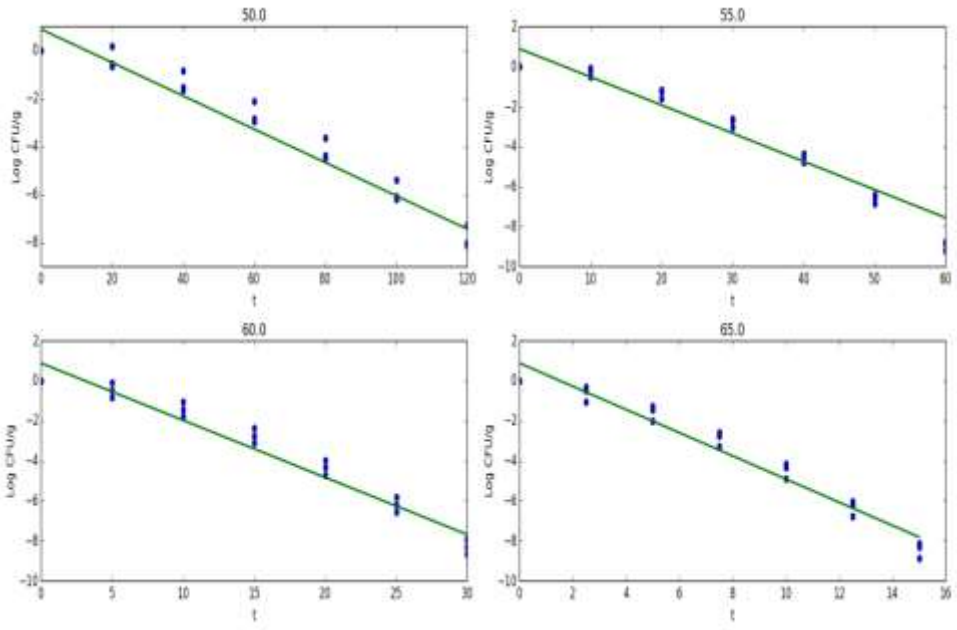

Figure 10. Multiple survival curves are analyzed simultaneously using the linear inactivation model

USDA Integrated Pathogen Modeling Program for Predictive Microbiology - IPMP Global Fit

Dr. Lihan Huang
USDA ARS Eastern Regional Research Center
600 E. Mermaid Lane, Wyndmoor, PA 19038
Email: lihan.huang@ars.usda.gov

Current time is: 2017-01-03 14:32:10.701000

Primary Model	Linear inactivation
Secondary Model	D/z
Degree of freedom	81
SSE	3.547e+01
MSE	4.378e-01
RMSE	6.617e-01
Residual Error Std	6.498e-01
AIC	-63.923

Estimated Parameters

Parameters	Value	Std-Error	t-value	p-value	L95CI	U95CI
logD0	4.23E+00	9.76E-02	43.38	7.76E-58	4.04E+00	4.43E+00
z	1.63E+01	4.40E-01	37.01	1.61E-52	1.54E+01	1.71E+01
y0	8.86E-01	1.97E-01	4.50	2.23E-05	4.94E-01	1.28E+00

Figure 11. Result of data analysis of **m**ultiple survival curves using the linear inactivation model.

2. Growth models

2.1 Primary models

2.1.1 No-lag phase model (Fang, Gurtler, and Huang, 2012; Fang, Liu, and Huang, 2013)

$$Y(t) = Y_0 + Y_{max} - ln[e^{Y_0} + (e^{Y_{max}} - e^{Y_0})e^{-\mu_{max}t}]$$

2.1.2 Huang Model (Huang, 2008, 2013)

$$Y(t) = Y_0 + Y_{max} - ln\{e^{Y_0} + [e^{Y_{max}} - e^{Y_0}]e^{-\mu_{max}B(t)}\}$$

$$B(t) = t + \frac{1}{4} ln \frac{1 + e^{-4(t-\lambda)}}{1 + e^{4\lambda}}$$

2.2 Secondary models

2.2.1 Full temperature range Ratkowsky square-root model (Ratkowsky et al., 1983)

$$\sqrt{\mu} = a(T - T_0)\left[1 - e^{b(T - T_{max})}\right]$$

2.2.2 Suboptimal Ratkowsky square-root model (Ratkowsky et al., 1982)

$$\sqrt{\mu} = a(T - T_0)$$

2.2.3 Full temperature range Huang square-root model (Huang, Hwang, and Phillips, 2011)

$$\sqrt{\mu} = a(T - T_{min})^{0.75}\left[1 - e^{b(T - T_{max})}\right]$$

2.2.4 Suboptimal Huang square-root model (Huang, Hwang, and Phillips, 2011)

$$\sqrt{\mu} = a(T - T_{min})^{0.75}$$

2.2.5 Cardinal parameters model (Rosso, Lobry, and Flandrois, 1993)

$$\mu_{max} = \frac{\mu_{opt}(T - T_{max})(T - T_{min})^2}{\left[(T_{opt} - T_{min})(T - T_{opt}) - (T_{opt} - T_{max})(T_{opt} + T_{min} - 2T)\right](T_{opt} - T_{min})}$$

2.2.6 Lag phase for Huang model

$$\lambda = \frac{e^A}{\mu_{max}^m}$$

Sometimes it may be difficult to determine both A and m during nonlinear regression. The errors of the estimates may be large if both A and m are included as parameters to estimate. If this situation arises, the software allows fixing A to a certain value if needed. If fixing A is needed, it is suggested you use the value of A estimated by the software when both A and m are included as

parameters. Fixing A remove it from the parameters to estimated, and may allow more accurate estimation of m for the lag phase. A can be both positive or negative.

2.3 Data analysis

For growth modeling, the users are expected to 1) make a choice of the primary models, 2) submit the choice by clicking the "Submit model" button, 3) select a secondary model from the pop-up window, 4) adjust the parameters in the secondary model window, and 5) submit the secondary model by clicking "Submit model" in the pop-up window. If a solution is found, the regression curves will be plotted with the raw data (Figure 12), and a report will be generated (Figure 13).

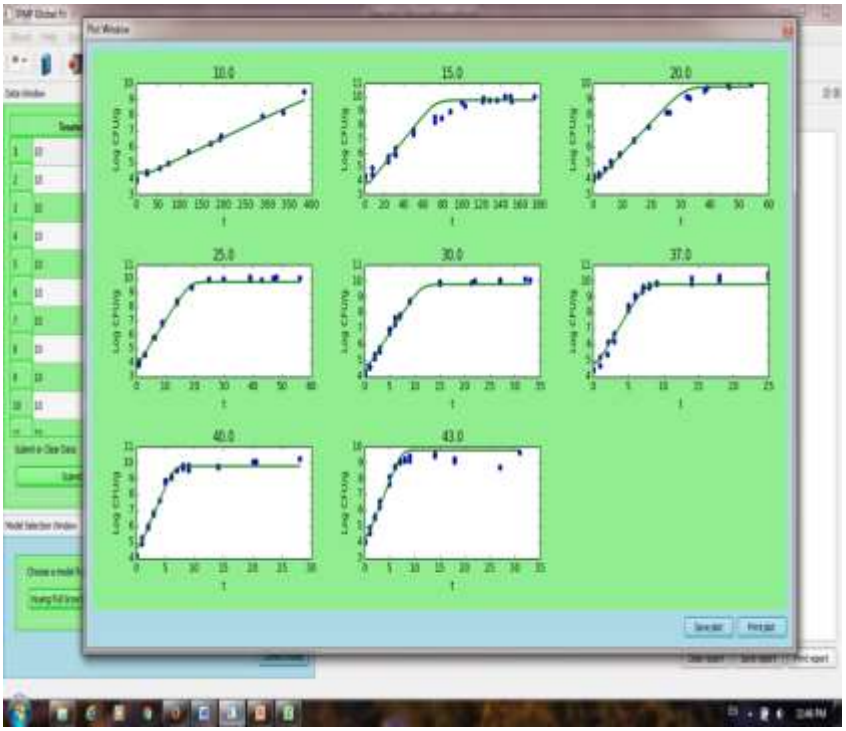

Figure 12. Regression curves of growth data.

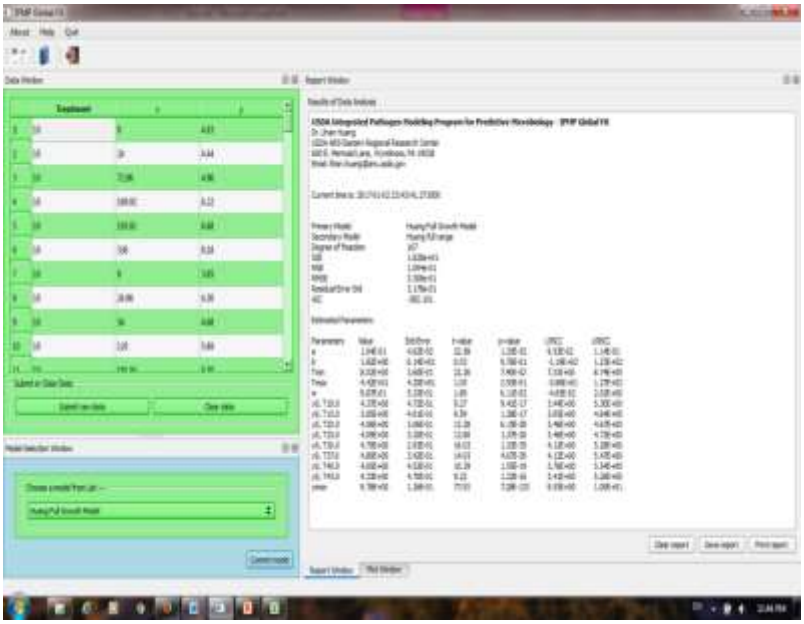

Figure 13. Result of data analysis for growth curves.

REFERENCES

Fang, T., Gurtler, J.B., and Huang, L. 2012. Growth kinetics and model comparison of *Cronobacter sakazakii* in reconstituted powdered infant formula. Journal of Food Science, 77: E247 – E255.

Fang, T., Liu, Y., and Huang, L. 2013. Growth kinetics of *Listeria monocytogenes* and spoilage microorganisms in fresh-cut cantaloupe. Food Microbiology, 34: 174 – 181.

Huang, L. 2008. Growth kinetics of *Listeria monocytogenes* in broth and beef frankfurters – determination of lag phase duration and exponential growth rate under isothermal conditions. Journal of Food Science, 73: E235 – 242.

Huang, L. 2013. Optimization of a new mathematical model for bacterial growth. Food Control, 32: 283 – 288.

Huang, L. 2015a. Growth of *Staphylococcus aureus* in cooked potato and potato salad - a one-step kinetic analysis. J. Food Science, 80(12): M2837-M2844.

Huang, L. 2015b. Direct construction of predictive models for describing growth of *Salmonella Enteritidis* in liquid eggs - a one-step approach. Food Control, 57: 76-81. 2015.

Huang, L. 2016. Mathematical modeling and validation of growth of *Salmonella Enteritidis* and background microorganisms in potato salad – one-step kinetic analysis and model development. Food Control, 68: 69-76. 2016.

Huang, L., Hwang, C., and Phillips, J.G. 2011a. Evaluating the effect of temperature on microbial growth rate - the Ratkowsky and a Belehrádek type models. Journal of Food Science, 76, M547-557.

Mafart, P., Couvert, O., Gaillard, S., and Leguerinel. 2002. On calculating sterility in thermal preservation methods: application of the Weibull frequency distribution model. International Journal of Food Microbiology, 72: 107 – 113.

Ratkowsky, D. A., J. Olley, T. A. McMeekin, and A. Ball. 1982. Relationship between temperature and growth rate of bacterial cultures. Journal of Bacteriology, 149:1-5.

Ratkowsky, D,A., Lowry, R.K., McMeekin, T.A., Stokes, A.N., and Chandler, R.E. 1983. Model for bacterial culture growth rate through the entire biokinetic temperature range. Journal of Bacteriology, 154: 1222–1226.

Rosso, L., Lobry, J.R., and Flandrois, J.P. 1993. An unexpected correlation between cardinal temperatures of microbial growth highlighted by a new model. Journal of Theoretical Biology, 162:447–63.